Gov 2000: 3. Multiple Random Variables

Matthew Blackwell

Fall 2016

Where are we? Where are we going?

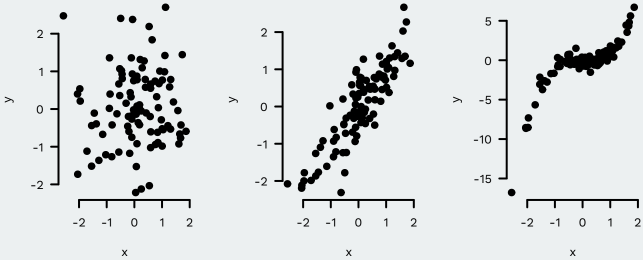
- Distributions of one variable: how to describe and summarize uncertainty about one variable.
- Today: distributions of multiple variables to describe relationships between variables.
- Later: use data to learn about probability distributions.

Why multiple random variables?

1. How do we summarize the relationship between two variables, X and Y?
2. What if we have many observations of the same variable, $X_1, X_2, \ldots, X_n$?

1/ Distributions of Multiple Random Variables

Joint distributions



- The joint distribution of two r.v.s, X and Y, describes what pairs of observations, (x, y) are more likely than others.
  - Settler mortality (X) and GDP per capita (Y) for the same country.
- Shape of the joint distribution now includes the relationship between X and Y

Discrete r.v.s

Joint probability mass function

The joint p.m.f. of a pair of discrete r.v.s, (X, Y) describes the probability of any pair of values:

$$f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$$

- Properties of a joint p.m.f.:
  - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)
  - $\sum_x \sum_y f_{X,Y}(x, y) = 1$ (something must happen)
  - $\sum_x$ is shorthand for sum over all possible values of X

Example: Gay marriage and gender

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.3	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:
  - Each cell is the probability of that combination, $f_{X,Y}(x, y)$
- Probability that we randomly select a woman who favors gay marriage?

$$f_{X,Y}(1, 1) = \mathbb{P}(X = 1, Y = 1) = 0.3$$

Marginal distributions

- Often need to figure out the distribution of just one of the r.v.s
  - Called the marginal distribution in this context.
- Computing marginals from the joint p.m.f.:

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x f_{X,Y}(x, y)$$

- Intuition: sum over the probability that $Y = y$ for all possible values of x
  - Works because these are mutually exclusive events that partition the space of X

Example: marginals for gay marriage

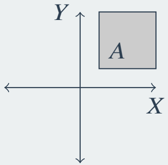
	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- What's the $f_Y(1) = \mathbb{P}(Y = 1)$?
  - Probability that a man favors gay marriage plus the probability that a woman favors gay marriage.

$$f_Y(1) = f_{X,Y}(1,1) + f_{X,Y}(0,1) = 0.3 + 0.22 = 0.52$$

- Works for all marginals.

Continuous r.v.s



- We will focus on getting the probability of being in some subset of the 2-dimensional plane.

Continuous joint p.d.f.

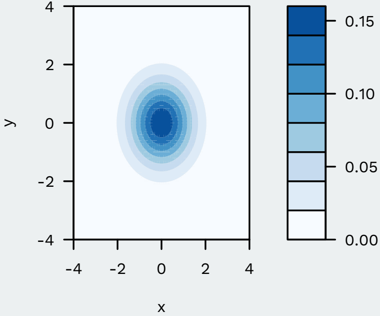
Continuous joint distribution

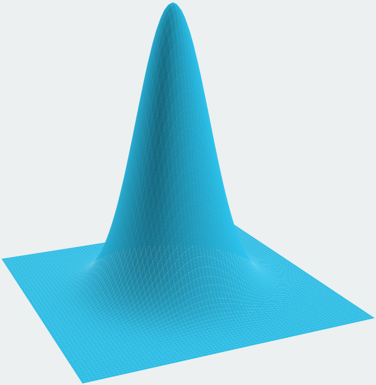
Two continuous r.v.s X and Y have a continuous joint distribution if there is a nonnegative function $f_{X,Y}(x, y)$ such that for any subset A of the xy-plane,

$$\mathbb{P}((X, Y) \in A) = \iint_{(x,y)\in A} f_{X,Y}(x, y)dxdy.$$

- $f_{X,Y}(x, y)$ is the joint probability density function.
- $\{(x, y) : f_{X,Y}(x, y) > 0\}$ is called the support of the distribution.
- Joint p.d.f. must meet the following conditions:
    1. $f_{X,Y}(x, y) \geq 0$ for all values of (x, y), (nonnegative)
    2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$, (probabilities “sum” to 1)
- $\mathbb{P}(X = x, Y = y) = 0$ for similar reasons as with single r.v.s.

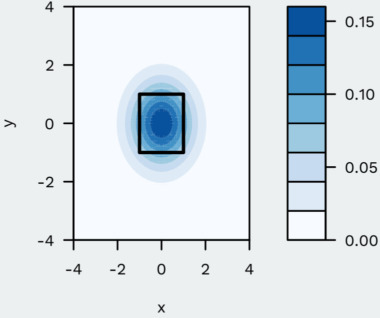
Joint densities are 3D

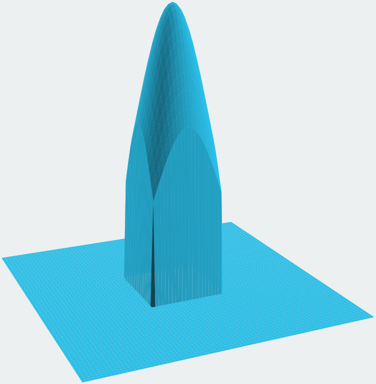




- X and Y axes are on the "floor," height is the value of $f_{X,Y}(x, y)$.
- Remember $f_{X,Y}(x, y) \neq \mathbb{P}(X = x, Y = y)$.

Probability = volume





- $\mathbb{P}((X,Y) \in A) = \iint_{(x,y)\in A} f_{X,Y}(x,y)dxdy$
- Probability $=$ volume above a specific region.

Working with joint p.d.f.s

- Suppose we have the following form of a joint p.d.f.:

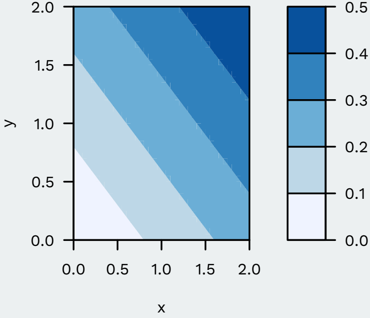
$$f_{X,Y}(x,y) = \begin{cases} c(x+y) & \text{for } 0 < x < 2 \text{ and } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

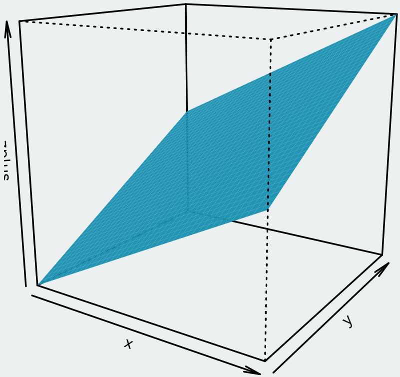
- What does c have to be for this to be a valid p.d.f.?

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dxdy \\ &= \int_0^2 \int_0^2 c(x+y) dxdy \\ &= c \int_0^2 \left(\frac{x^2}{2} + xy \right) \bigg|_{x=0}^{x=2} dy \\ &= c \int_0^2 (2+2y) dy \\ &= \left(2cy + cy^2\right)\Big|_0^2 = 8c \end{aligned}$$

- Thus to be a valid p.d.f., $c = 1/8$

Example continuous distribution





$$f_{X,Y}(x,y) = \begin{cases} (x+y)/8 & \text{for } 0 < x < 2 \text{ and } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

Continuous marginal distributions

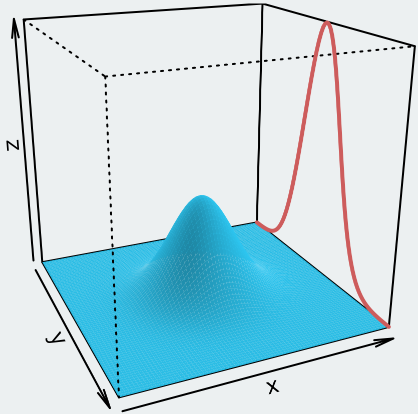
- We can recover the marginal PDF of one of the variables by integrating over the distribution of the other variable:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

- Works for either variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

Visualizing continuous marginals



- Marginal integrates (sums, basically) over other r.v.: $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$
- Pile up/flatten all of the joint density onto a single dimension.

Deriving continuous marginals

$$f_{X,Y}(x,y) = \begin{cases} (x+y)/8 & \text{for } 0 < x < 2 \text{ and } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

- Let's calculate the marginals for this p.d.f.:

$$\begin{aligned} f_X(x) &= \int_0^2 \frac{1}{8}(x+y)dy \\ &= \left(\frac{xy}{8} + \frac{y^2}{16} \right) \Bigg|_{y=0}^{y=2} \\ &= \frac{x}{4} + \frac{1}{4} = \frac{x+1}{4} \end{aligned}$$

- By symmetry we have the same for y:

$$f_Y(y) = (y+1)/4$$

Joint c.d.f.s

Joint cumulative distribution function

For two r.v.s X and Y, the joint cumulative distribution function or joint c.d.f. $F_{X,Y}(x,y)$ is a function such that for finite values x and y,

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

- Deriving p.d.f. from c.d.f.: $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$
- Deriving c.d.f. from p.d.f: $F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(r,s) dr ds$

2/ Properties of Joint Distributions

Properties of joint distributions

- Single r.v.: summarized $f_X(x)$ with $\mathbb{E}[X]$ and $\mathbb{V}[X]$
- With 2 r.v.s, we can additionally measure how strong the dependence is between the variables.
- First: expectations over joint distributions and independence

Expectations over multiple r.v.s

- 2-d LOTUS: take expectations over the joint distribution.
- With discrete X and Y:

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) f_{X,Y}(x, y)$$

- With continuous X and Y:

$$\mathbb{E}[g(X, Y)] = \int_x \int_y g(x, y) f_{X,Y}(x, y) dxdy$$

- Marginal expectations:

$$\mathbb{E}[Y] = \sum_x \sum_y y f_{X,Y}(x, y)$$

- Example: expectation of the product:

$$\mathbb{E}[XY] = \sum_x \sum_y xy f_{X,Y}(x, y)$$

Marginal expectations from joint

$$f_{X,Y}(x,y) = \begin{cases} (x+y)/8 & \text{for } 0 < x < 2 \text{ and } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

- Marginal expectation of Y:

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^2 \int_0^2 y \frac{1}{8}(x+y)dxdy \\ &= \int_0^2 y \int_0^2 \frac{1}{8}(x+y)dxdy \\ &= \int_0^2 y \frac{1}{4}(y+1)dy \\ &= \left(\frac{y^3}{12} + \frac{y^2}{8} \right)\Big|_0^2 \\ &= \frac{2}{3} + \frac{1}{2} = \frac{7}{6} \end{aligned}$$

- By symmetry, $\mathbb{E}[X] = \mathbb{E}[Y] = 7/6$

Independence

Independence

Two r.v.s Y and X are independent (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

- Knowing the value of X gives us no information about the value of Y.
- If X and Y are independent, then:
  - $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (joint is the product of marginals)
  - $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
  - $h(X) \perp\!\!\!\perp g(Y)$ for any functions $h()$ and $g()$ (functions of independent r.v.s are independent)

Key properties of independent r.v.s

- Theorem If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy f_{X,Y}(x,y) \\
&= \sum_x \sum_y xy f_X(x) f_Y(y) \\
&= \left(\sum_x x f_X(x)\right)\left(\sum_y y f_Y(y)\right) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Why independence?

- Independence assumptions are everywhere in theoretical and applied statistics.
  - Each response in a poll is considered independent of all other responses.
  - In a randomized control trial, treatment assignment is independent of background characteristics.
- Lack of independence is a blessing or a curse:
  - Two variables not independent $\rightsquigarrow$ potentially interesting relationship.
  - In observational studies, treatment assignment is usually not independent of background characteristics.

Covariance

- If two variables are not independent, how do we measure the strength of their dependence?
  - Covariance
  - Correlation
- Covariance: how do two r.v.s vary together?
  - How often do high values of X occur with high values of Y?

Defining covariance

- If two variables are not independent, how do we measure the strength of their dependence?

Covariance

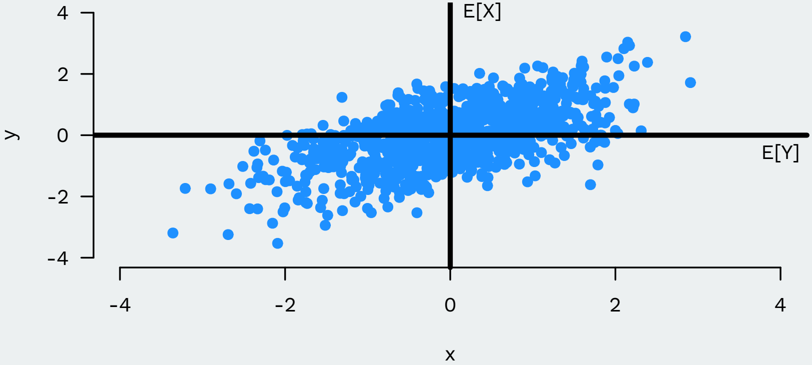
The covariance between two r.v.s, X and Y is defined as:

$$\text{Cov}[X, Y] = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big]$$

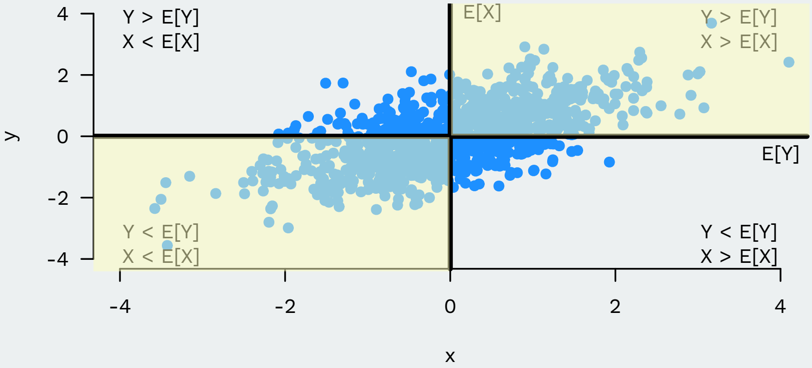
- How often do high values of X occur with high values of Y?
- Properties of covariances:
  - $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
  - If $X \perp\!\!\!\perp Y$,

$$\begin{aligned}\text{Cov}[X, Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0\end{aligned}$$

Covariance intuition

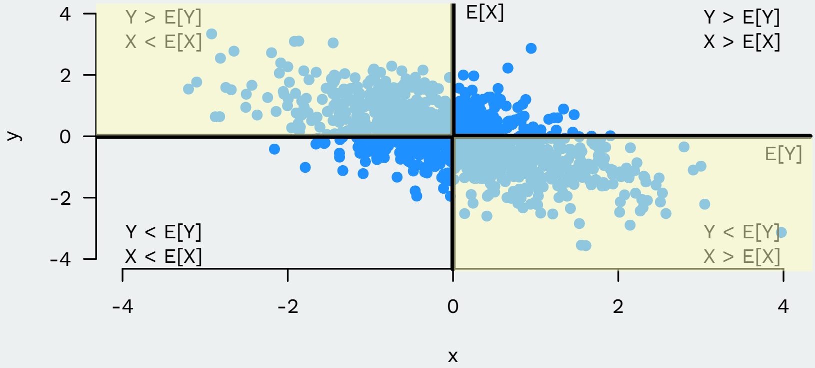


Covariance intuition



- Large values of X tend to occur with large values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{pos. num.}) \times (\text{pos. num}) = +$
- Small values of X tend to occur with small values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{neg. num.}) \times (\text{neg. num}) = +$
- If these dominate $\rightsquigarrow$ positive covariance.

Covariance intuition



- Large values of X tend to occur with small values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{pos. num.}) \times (\text{neg. num}) = -$
- Small values of X tend to occur with large values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{neg. num.}) \times (\text{pos. num}) = -$
- If these dominate $\rightsquigarrow$ negative covariance.

Covariance from joint p.d.f.

- Using our running example of $f_{X,Y}(x, y) = (x + y)/8$
- From earlier: $\mathbb{E}[X] = \mathbb{E}[Y] = 7/6$
- Expectation of the product:

$$\begin{aligned}
\mathbb{E}[XY] &= \int_0^2 \int_0^2 xy\frac{1}{8}(x + y)dxdy \\
&= \int_0^2 \int_0^2 \frac{1}{8}(x^2y + xy^2)dxdy \\
&= \int_0^2 \left(\frac{x^3y}{24} + \frac{x^2y^2}{16}\right)\Bigg|_{x=0}^{x=2} dy \\
&= \int_0^2 \left(\frac{y}{3} + \frac{y^2}{4}\right) dy \\
&= \left(\frac{y^2}{6} + \frac{y^3}{12}\right)\Bigg|_0^2 = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}
\end{aligned}$$

- Covariance:

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{4}{3} - \left(\frac{7}{6}\right)^2 = -\frac{1}{36}$$

Zero covariance doesn't imply independence

- We saw that $X \perp\!\!\!\perp Y \rightsquigarrow \text{Cov}[X, Y] = 0$.
- Does $\text{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$? No!
- Counterexample: $X \in \{-1, 0, 1\}$ with equal probability and $Y = X^2$.
- Covariance is a measure of linear dependence, so it can miss non-linear dependence.

Properties of variances and covariances

- Properties of covariances:

1. $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$.
2. $\text{Cov}[X, X] = \mathbb{V}[X]$

- Properties of variances that we can state now that we know covariance:

1. $\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] + 2ab\text{Cov}[X, Y]$
2. If X and Y independent, $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Using properties of covariance

- Rescale our running example: $Z = 2X$, $W = 2Y$.
- What's the covariance of (Z, W)?
  - Ugh, let's avoid more integrals.
- Use properties of covariances:

$$\text{Cov}[Z, W] = \text{Cov}[2X, 2Y] = 2 \times 2 \times \text{Cov}[X, Y] = -\frac{1}{9}$$

Correlation

- Covariance is not scale-free: $\text{Cov}[2X, Y] = 2\text{Cov}[X, Y]$
  - $\rightsquigarrow$ hard to compare covriances across different r.v.s
  - Is a relationship stronger? Or just do to rescaling?
- Correlation is a scale-free measure of linear dependence.

Correlation

The correlation between two r.v.s X and Y is defined as:

$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}}$$

- Covariance after dividing out the scales of the respective variables.
- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - if $|\rho(X, Y)| = 1$, then X and Y are perfectly correlated with a deterministic linear relationship: $Y = a + bX$.

3/ Conditional Distributions

Conditional distributions

- Conditional distribution: distribution of Y if we know $X = x$.

Conditional probability mass function

The conditional probability mass function or conditional p.m.f. of Y conditional on X is

$$f_{Y|X}(y|x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

- Intuitive definition:

$$f_{Y|X}(y|x) = \frac{\text{Probability that } X = x \text{ and } Y = y}{\text{Probability that } X = x}$$

- This is a valid univariate probability distribution!
  - $f_{Y|X}(y|x) \geq 0$ and $\sum_y f_{Y|X}(y|x) = 1$
- If $X \perp\!\!\!\perp Y$ then $f_{Y|X}(y|x) = f_Y(y)$ (conditional is the marginal)

Example: conditionals for gay marriage

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on being a man?

$$f_{Y|X}(y = 1|x = 0) = \frac{\mathbb{P}(X = 0, Y = 1)}{\mathbb{P}(X = 0)} = \frac{0.22}{0.22 + 0.27} = 0.44$$

Example: conditionals for gay marriage

- Two values of $X \rightsquigarrow$ two univariate conditional distribution of Y

Continuous conditional distributions

Conditional probability density function

The conditional p.d.f. of a continuous random variable is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

assuming that $f_X(x) > 0$.

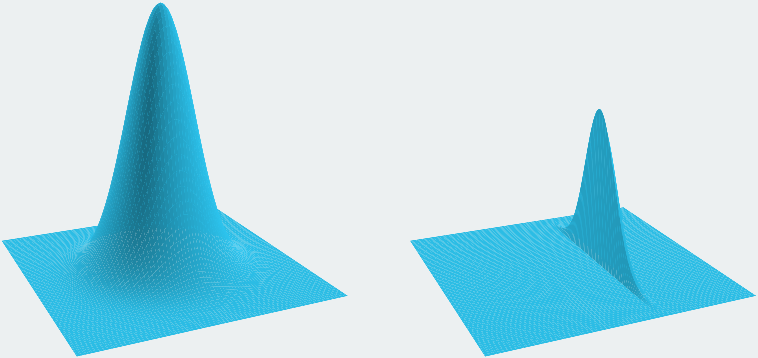
- Implies

$$\mathbb{P}(a < Y < b|X = x) = \int_a^b f_{Y|X}(y|x)dy.$$

- Based on the definition of the conditional p.m.f./p.d.f., we have the following factorization:

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$$

Conditional distributions as slices



- $f_{Y|X}(y|x_0)$ is the conditional p.d.f. of Y when $X = x_0$
- $f_{Y|X}(y|x_0)$ is proportional to joint p.d.f. along x_0: $f_{X,Y}(y, x_0)$
- Normalize by dividing by $f_X(x_0)$ to ensure proper p.d.f.

Continuous conditional example

- Using our running example of $f_{X,Y}(x, y) = (x + y)/8$
- Earlier we calculated $f_X(x) = (x + 1)/4$
- Calculate conditional:

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\
&= \frac{(x + y)/8}{(x + 1)/4} \\
&= \frac{x + y}{2(x + 1)}
\end{aligned}
$$

- Remember the limits: $0 < y < 2$, 0 otherwise

Conditional Independence

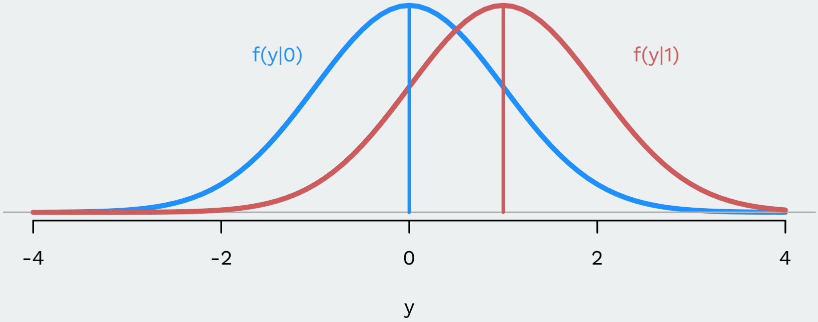
Conditional independence

Two r.v.s X and Y are conditionally independent given Z (written $X \perp\!\!\!\perp Y|Z$) if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z).$$

- X and Y are independent within levels of Z.
- Massively important for regression, causal inference.
- Example:
  - X = swimming accidents, Y = number of ice cream cones sold.
  - In general, dependent.
  - Conditional on Z = temperature, independent.

Summarizing conditional distributions



- Conditional distributions are also univariate distribution and so we can summarize them with its mean and variance.
- Gives us insight into a key question:
  - How does the mean of Y change as we change X?

Defining condition expectations

Conditional expectation

The conditional expectation of Y conditional on $X = x$ is:

$$\mathbb{E}[Y|X = x] = \begin{cases} \sum_{y} y f_{Y|X}(y|x) & \text{discrete } Y \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy & \text{continuous } Y \end{cases}$$

- Intuition: exactly the same definition of the expected value with $f_{Y|X}(y|x)$ in place of $f_Y(y)$
- The expected value of the (univariate) conditional distribution.
- This is a function of x!

Calculating conditional expectations

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- What's the conditional expectation of support for gay marriage Y given someone is a man $X = 0$?

$$\begin{aligned}\mathbb{E}[Y|X = 0] &= \sum_y y f_{Y|X}(y|x = 0) \\ &= 0 \times f(y = 0|x = 0) + 1 \times f(y = 1|x = 0) \\ &= 1 \times \frac{0.22}{0.22 + 0.27} \\ &= 0.44\end{aligned}$$

Conditional expectations are random variables

- For a particular x, $\mathbb{E}[Y|X = x]$ is a number.
- But X takes on many possible values with uncertainty
  $\rightsquigarrow$ $\mathbb{E}[Y|X]$ takes on many possible values with uncertainty.
- $\rightsquigarrow$ Conditional expectations are random variables!
- Binary X:

$$\mathbb{E}[Y|X] = \begin{cases} \mathbb{E}[Y|X = 0] & \text{with prob. } \mathbb{P}(X = 0) \\ \mathbb{E}[Y|X = 1] & \text{with prob. } \mathbb{P}(X = 1) \end{cases}$$

- Has an expectation, $\mathbb{E}[\mathbb{E}[Y|X]]$, and a variance, $\mathbb{V}[\mathbb{E}[Y|X]]$.

Law of iterated expectations

- Average/mean of the conditional expectations: $\mathbb{E}[\mathbb{E}[Y|X]]$.
  - Can we connect this to the marginal (overall) expectation?
- **Theorem** (The Law of Iterated Expectations): If the expectation exist and for discrete X,

$$\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y|X]\right] = \sum_x \mathbb{E}[Y|X = x] f_X(x)$$

Example: law of iterated expectations

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	1

- $\mathbb{E}[Y|X=1] = 0.59$ and $\mathbb{E}[Y|X=0] = 0.44$.
- $f_X(1) = 0.51$ (females) and $f_X(0) = 0.49$ (males).
- Plug into the iterated expectations:

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \mathbb{E}[Y|X=0]f_X(0) + \mathbb{E}[Y|X=1]f_X(1) \\ &= 0.44 \times 0.49 + 0.59 \times 0.51 \\ &= 0.52 = \mathbb{E}[Y]\end{aligned}$$

Properties of conditional expectations

1. $\mathbb{E}[c(X)|X] = c(X)$ for any function $c(X)$.
    - Example: $\mathbb{E}[X^2|X] = X^2$ (If we know X, then we also know X^2)
2. If X and Y are independent r.v.s, then

$$\mathbb{E}[Y|X = x] = \mathbb{E}[Y].$$

3. If $X \perp\!\!\!\perp Y|Z$, then

$$\mathbb{E}[Y|X = x, Z = z] = \mathbb{E}[Y|Z = z].$$

Conditional Variance

Conditional expectation

The conditional variance of a Y given $X = x$ is defined as:

$$\mathbb{V}[Y|X = x] = \mathbb{E}\left[(Y - \mathbb{E}[Y|X = x])^2|X = x\right]$$

- Conditional variance describes the spread of the conditional distribution around the conditional expectation.
- Usual variance formula applied to conditional distribution.
- Using LOTUS:
  - Discrete Y:

$$\mathbb{V}[Y|X = x] = \sum_y (y - \mathbb{E}[Y|X = x])^2 f_{Y|X}(y|x)$$

  - Continuous Y:

$$\mathbb{V}[Y|X = x] = \int_y (y - \mathbb{E}[Y|X = x])^2 f_{Y|X}(y|x) dy$$

Conditional variance is a random variable

- Again, $\mathbb{V}[Y|X]$ is a random variable and a function of X, just like $\mathbb{E}[Y|X]$. With a binary X:

$$\mathbb{V}[Y|X] = \begin{cases} \mathbb{V}[Y|X=0] & \text{with prob. } \mathbb{P}(X=0) \\ \mathbb{V}[Y|X=1] & \text{with prob. } \mathbb{P}(X=1) \end{cases}$$

Law of total variance

- We can also relate the marginal variance to the conditional variance and the conditional expectation.
- **Theorem** (Law of Total Variance/EVE's law):

$$\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y|X]] + \mathbb{V}[\mathbb{E}[Y|X]]$$

- The total variance can be decomposed into:
  1. the average of the within group variance ($\mathbb{E}[\mathbb{V}[Y|X]]$) and
  2. how much the average varies between groups ($\mathbb{V}[\mathbb{E}[Y|X]]$).

4/ Wrap-up

Review

- Multiple r.v.s require joint p.m.f.s and joint p.d.f.s
- Multiple r.v.s can have distributions that exhibit dependence as measured by covariance and correlation.
- The conditional expectation of one variable on the other is an important quantity that we'll see over and over again.